doi: 10.17221/8240-CJAS

Contribution of domestic and Interbull records to reliabilities of single-step genomic breeding values in dairy cattle

J. Bauer, J. Přibyl, L. Vostrý

Institute of Animal Science, Prague-Uhříněves, Czech Republic

ABSTRACT: The method of approximating reliabilities of genomic breeding values in the single-step genomic BLUP evaluation procedure of Misztal et al. (2013) was used to evaluate the increase in reliability of breeding values for milk production in dairy cattle brought about by the inclusion of genomic data. Three strategies for approximation of reliabilities were compared: using only domestic records from performance testing of cows in the Czech Holstein dairy cattle population, using the same records in combination with Interbull breeding values of sires expressed as deregressed proofs, and using only the Interbull breeding values of sires expressed as deregressed proofs. The highest average reliability of genomic breeding values was achieved by the strategy using both domestic and Interbull data, for which the approximated reliabilities of genotyped bulls increased by 0.063. This general increase in reliability of genomic breeding values was small due to the small number of reference bulls available for the study. The overall increase in reliabilities for the entire population of dairy cattle was low but detectable. That modest increase was partially dependent on the unfavourable ratio of the number of genotyped bulls to the size of the analyzed population. Inclusion of Interbull data dramatically increased the benefits of genotyping in our test case – a relatively small population with substantial genetic contributions of foreign genes.

Keywords: data combining; genomic selection; reliability; single-step GBLUP

INTRODUCTION

Prediction of genomic breeding values is a promising procedure of how to increase accuracy of genetic evaluation in cattle. It could be of a special benefit namely for young animals lacking performance data. Two types of procedures can be used to predict genomic breeding values: the multi-step method of VanRaden (2008) and the single-step method of Misztal et al. (2009) and Christensen and Lund (2010).

In both methodologies, numerical reliabilities of predicted individual genomic breeding values (GEBV) are required to optimize the selection of parents for the next generation. These reliabilities can be calculated by the inversion of the left hand side of a BLUP system of equations (VanRaden 2008), but frequently this is not feasible due to large population sizes and massive computational requirements. For this reason, methods to approximate reliabilities of GEBV were developed by Szyda et al. (2011) for the multi-step method and by Misztal et al. (2013) for the single-step method cited above.

Because the single-step prediction of genomic breeding values is undergoing a successful development in the Czech Republic (Pribyl et al. 2012, 2013), methods to calculate reliabilities of predicted genomic breeding values need to be established as well. Therefore, the aim of this study was to compare the efficacy of the Misztal et al. (2013) method for the approximation of reliability of single-step genomic breeding values in Czech dairy cattle when different sources of information were used in the evaluation.

Supported by the Ministry of Agriculture of the Czech Republic (Projects No. QI111A167 and No. MZE RO0714).

MATERIAL AND METHODS

Data from 861 429 first lactations of Czech Holstein cows recorded during the calving years 1991–2006 were used in this study. The 4-generation pedigree included 1 868 252 animals. In total, 1479 Holstein bulls were genotyped using the Illumina BovineSNP50 Beadchip V2; 801 animals were used in the reference population. Data were similar to those used by Pribyl et al. (2013). The evaluated trait was the total first lactation milk yield.

For reliability approximations, a heritability estimate of 0.25 taken from previous analyses (data not shown) was used. For inclusion of Interbull data, March 2012 breeding values were utilized.

Our analysis for the approximation of reliability of single-step genomic breeding values is based on Misztal et al. (2013), and the procedure can be described as a sequence of consecutive steps:

(1) to approximate the reliabilities for the entire population using the iterative approach of Misztal et al. (1993)

(2) to convert the reliabilities to the effective number of records for genotyped animals by the formula:

$$d_i = \alpha\ (1/\text{reliability} - 1)$$

where:

α = ratio of error variance to additive genetic variance

(3) to update the reliabilities only for the genotyped animals by $\mathbf{Q}^{-1}$, the inversion of the sum of the effective number of records and the relationship matrix calculated by the formula of Misztal et al. (2013):

$$\mathbf{Q}^{-1} = [\mathbf{D} + (\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}^{-1}_{22})\ \alpha]^{-1}$$

where:

$\mathbf{D}$ = contribution of records and pedigrees to reliability
$\mathbf{I}$ = identity matrix
$\mathbf{G}^{-1}$ = inverse of genomic relationship matrix
$\mathbf{A}^{-1}_{22}$ = inverse of section of relationship matrix that contains pedigree-based relationship information on genotyped animals

(4) to calculate reliabilities (r_i^2) for the genotyped animals:

$$r_i^2 = 1 - \alpha\ q_{ii}$$

where:

q_{ii} = diagonal elements of $\mathbf{Q}^{-1}$ matrix

(5) To add the contribution from genotyping to the reliability of the non-genotyped animals. This step was completed using a procedure similar to that described in step (1). The reliabilities of the genotyped bulls were held constant to avoid double counting the contribution of the relationships among these bulls and to avoid changing the already fully conveyed values.

A relaxation parameter was applied to reduce oscillation between iterations and to improve convergence in the first and fifth steps of the method. The relaxation parameter ω was set to 0.5 and effective number of records for animal i at the beginning of iteration $n + 2$ was as follows:

$$d_i^{n+2} = \omega\ (d_i^{n} + d_i^{n+1})$$

The entry values for the $\mathbf{G}$ and $\mathbf{A}_{22}$ matrices were acquired from the calculations of Pribyl et al. (2013). The $\mathbf{G}$ matrix had already been scaled to the mean of the diagonal elements equal to 1 by the procedure of Forni et al. (2011) and shifted to $\mathbf{A}_{22}$ matrix values as proposed by Vitezica et al. (2011).

The information sources for evaluation used in the three strategies were (A) domestic (Czech population) records from performance testing, (B) domestic records from performance testing and Interbull breeding values (EBVs) expressed as deregressed proofs, and (C) Interbull EBVs expressed as deregressed proofs.

The combination of the effective number of records for animals given by the domestic records and Interbull EBVs was carried out prior to the iterative part of step (1) by the following formula:

$$d_i^b = \sum_{p=1}^{N} \left(w_{ip} - \frac{w_{ip}^2}{\sum_{j=1}^{n_p} w_{jp}} \right)$$

where:

d_i^b = effective number of records after addition of other sources of information for animal i
w_{ip} = weight of records of animal i in a contemporary group p
n_p = number of animals in contemporary group p
N = number of contemporary groups p

The individual weights of records of the animals in contemporary groups were calculated as:

$$w_{ip} = \alpha\ \frac{r_i^2}{1 - r_i^2}$$

where:

α = ratio of residual to additive genetic variance
r_i^2 = reliability of breeding value provided by additional source of information, in our case reliabilities of Interbull EBVs

doi: 10.17221/8240-CJAS

Table 1. Descriptive statistics for reliabilities of GEBV from domestic production records (strategy "A")

	Approximate reliability	*n*	Mean	SD	Median	Min	Max
Entire population	EBV	1 651 651	0.325	0.135	0.338	0	0.998
Genotyped bulls	EBV	1 479	0.583	0.303	0.691	0.09	0.998
Genotyped bulls	GEBV	1 479	0.590	0.307	0.590	0	1
Δ Genotyped bulls	GEBV	1 479	0.006	0.079	0.003	–0.71	0.76
Δ Entire population	GEBV	1 651 651	0	0.008	0	–0.71	0.76

EBV = BLUP breeding values, GEBV = single-step genomic BLUP breeding values, Δ = difference from approximated reliability of EBV

All Interbull EBVs were taken into account as one contemporary group of animals. The sources of information (i.e. domestic records and Interbull EBVs) were checked to avoid double counting of the same information on individual animals.

The significance of differences of average reliability estimates between EBVs and GEBV was tested by *t*-tests for the entire population and for the genotyped bulls.

RESULTS AND DISCUSSION

Descriptive statistics for approximate reliabilities of milk yield GEBV estimated by the three strategies are shown in Tables 1–3. Approximations of the EBV reliabilities were computed by the first step of the Misztal et al. (2013) method (without utilization of genomic information in step (1) of the method) and for the GEBV reliabilities calculation the entire method was applied.

Although inclusion of genomic data in approximation of reliabilities of genomic breeding values on the basis of domestic performance testing data (i.e. strategy "A") was beneficial, the largest increase in approximate reliability of breeding values in comparison to the reliabilities of conventional breeding values was found in strategy "B", which incorporated domestic records and Interbull EBVs. The increase in average reliability of genotyped bulls was equal to 0.063. The "C" strategy followed, with an increase of 0.051. Using domestic records as the only data source resulted in the reliability increase by just 0.006.

These results are influenced by the fact that the Czech dairy cattle population is small and moreover strongly influenced by the import of foreign genes (Pribyl et al. 2012). Bulls used in breeding are often more closely related to international than to domestic populations. So although the domestic records are useful for genomic evaluation (as seen in strategy "A"), such information is more efficiently utilized in combination with Interbull EBVs in strategy "B". In that case, domestic records contribute significantly to the evaluation of genotyped bulls as well as to the entire population. This interpretation is supported by the difference in results between strategy "B" and strategy "C", in which only Interbull EBVs were used. Use of Interbull data for genomic evaluation can be beneficial for other countries in a similar situation.

The overall increase in the reliability of breeding values of genotyped bulls resulting from utilization of genomic information was small, mainly because of the small number (801) of reference bulls and because the increase in the reliability of breeding values by addition of genomic information is approximately linearly dependent upon the number

Table 2. Descriptive statistics for reliabilities of GEBV from domestic records and Interbull EBVs (strategy "B")

	Approximate reliability	*n*	Mean	SD	Median	Min	Max
Entire population	EBV	1 868 252	0.342	0.168	0.345	0	0.999
Genotyped bulls	EBV	1 479	0.623	0.289	0.744	0.129	0.999
Genotyped bulls	GEBV	1 479	0.686	0.236	0.773	0	1
Δ Genotyped bulls	GEBV	1 479	0.063	0.099	0.015	–0.421	0.767
Δ Entire population	GEBV	1 868 252	0.0005	0.012	0	–0.421	0.767

EBV = BLUP breeding values, GEBV = single-step genomic BLUP breeding values, Δ = difference to approximated reliability of EBV

doi: 10.17221/8240-CJAS

Table 3. Descriptive statistics for reliabilities of GEBV from Interbull EBVs (strategy "C")

	Approximate reliability	*n*	Mean	SD	Median	Min	Max
Entire population	EBV	258 644	0.420	0.311	0.424	0	0.999
Genotyped bulls	EBV	1 479	0.634	0.313	0.843	0.076	0.999
Genotyped bulls	GEBV	1 479	0.685	0.264	0.853	0	1
Δ Genotyped bulls	GEBV	1 479	0.051	0.072	0.051	–0.423	0.693
Δ Entire population	GEBV	258 644	0.0005	0.010	0	–0.40	0.750

EBV = BLUP breeding values, GEBV = single-step genomic BLUP breeding values, Δ = difference to approximated reliability of EBV

of reference bulls (VanRaden et al. 2009). While the increase in average reliability was also noticeable in non-genotyped animals, the small number of genotyped bulls relative to the size of the entire population resulted in only a small increase.

All differences between the average reliabilities of BLUP and Genomic BLUP breeding values were statistically significant ($P < 0.0001$) both for the entire population and for the genotyped bulls, with the exception of genotyped bulls in strategy "A" ($P = 0.0021$).

Misztal et al. (2013) reported that their computation method to approximate reliability for GEBVs could result in inflated values compared to reliabilities derived from such multi-step approaches as described by VanRaden et al. (2009). As noted by Misztal et al. (2013), inflated values can also result from the approximations and assumptions used in the multi-step method or from the reduction of genetic relationships over generations.

The general approach of pooling data is beneficial in that it increases the size of the reference population. Reliabilities based upon combined data are then higher. This was the rationale for combining the US and Canadian data into one genomic prediction. Likewise, the multinational Eurogenomics project was initiated because individual national genomic predictions were less accurate than one common evaluation (Lund et al. 2011). Our results showed the same trend. However, although the inclusion of animals from another population can be beneficial, a more important factor influencing the reliability of evaluation is the genetic relationship between the evaluated population and the reference population. The best results have been reported when the sires of young genotyped animals were themselves involved in the reference population (Lund et al. 2009).

As reported by VanRaden et al. (2009), the reliability of genomic evaluation is approximately linearly dependent upon the number of bulls in the reference population. On the other hand, with the increasing number of generations of ancestors included in the reference population the prediction reliability decreases (Pszczola et al. 2012). To prevent this decrease, Habier et al. (2010) proposed constant updates of reference animals from more recent generations.

The strong genetic connection between the genetic evaluation of Czech national data and of Interbull information, both being regularly updated, should justify the common evaluation. Nevertheless the combination can be problematic because the Interbull database includes Czech domestic data in its input and some information might be double counted. We took this situation into account and carefully filtered the data in our analysis, but this could be difficult to manage in routine genomic evaluation. Such difficulties could be outweighed by the positive influence of a smaller number of mutual relationships in Interbull than in domestic data, because having more unrelated animals in the reference population improves the accuracy of genomic evaluation (Pszczola et al. 2012). The benefit of Interbull data can be substantial, especially for less heritable traits for which there is generally a higher requirement for performance data or pseudo data (De Roos et al. 2009). In contrast to the increase of GEBV reliability with the size of reference population, the increase in the number of SNP markers from 40k to 600k typically improves the genomic prediction accuracy just to a limited extent (Pryce et al. 2012).

CONCLUSION

The reliability approximations of the single-step genomic BLUP GEBVs as described by Misztal et al. (2013) have increased by incorporating genotypic information into the evaluation of animals.

Interbull data can strongly impact the reliability of genetic evaluation in small populations with a substantial influence of foreign genes because of better utilization of genomic information.

REFERENCES

Christensen O.F., Lund M.S. (2010): Genomic prediction when some animals are not genotyped. Genetics Selection Evolution, 42: 2.

De Roos A.P.W., Hayes B.J., Goddard M.E. (2009): Reliability of genomic predictions across multiple populations. Genetics, 183, 1545–1553.

Forni S., Aguilar I., Misztal I. (2011): Different genomic relationship matrices for single step analysis using phenotypic, pedigree and genomic information. Genetics Selection Evolution, 43: 1.

Habier D., Tetens J., Seefried F.R., Lichtner P., Thaller G. (2010): The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution, 42: 5.

Lund M.S., Su G., Nielsen U.S., Aamand G.P. (2009): Relation between accuracies of genomic predictions and ancestral links to the training data. Interbull Bulletin, 40, 162–166.

Lund M.S., de Roos A.P., de Vries A.G., Druet T., Ducrocq V., Fritz S., Su G. (2011): A common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution, 43: 43.

Misztal I., Legarra A., Short T.H. (1993): Implementation of single- and multiple-trait animal models for genetic evaluation of Holstein type traits. Journal of Dairy Science, 76, 1421–1432.

Misztal I., Legarra A., Aguilar I. (2009): Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science, 92, 4648–4655.

Misztal I., Tsuruta S., Aguilar I., Leggara A., Van Raden P.M., Lawlor T.J. (2013): Methods to approximate reliabilities in single-step genomic evaluation. Journal of Dairy Science, 96, 647–654.

Pribyl J., Haman J., Kott T., Pribylova J., Simeckova M., Vostry L., Zavadilova L., Cermak V., Ruzicka Z., Splichal J., Verner M., Motycka J., Vondrasek L. (2012): Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes. Czech Journal of Animal Science, 57, 151–159.

Pribyl J., Madsen P., Bauer J., Pribylova J., Simeckova M., Vostry L., Zavadilova L. (2013): Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. Journal of Dairy Science, 96, 1865–1873.

Pryce J.E., Arias J., Bowman P.J., Davis S.R., Macdonald K.A., Waghorn G.C., Hayes B.J. (2012): Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. Journal of Dairy Science, 95, 2108–2119.

Pszczola M., Strabel T., Mulder H.A., Calus M.P.L. (2012): Reliability of direct genomic values for animals with different relationships within and to the reference population. Journal of Dairy Science, 95, 389–400.

Szyda J., Zarnecki A., Suchocki T., Kaminski S. (2011): Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. Journal of Applied Genetics, 52, 363–366.

VanRaden P.M. (2008): Efficient methods to compute genomic predictions. Journal of Dairy Science, 91, 4414–4423.

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., Schenkel F.S. (2009): Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 29, 16–24.

Vitezica Z.G., Aguilar I., Misztal I., Legarra A. (2011): Bias in genomic predictions for populations under selection. Genetic Research (Cambridge), 93, 357–366.

Received: 2014–06–09
Accepted after corrections: 2014–11–05

Corresponding Author

Jiří Bauer, Ph.D., Institute of Animal Science, Přátelství 815, 104 01 Prague 10-Uhříněves, Czech Republic
Phone: +420 267 009 574, e-mail: bauer.jiri@vuzv.cz